Bias in data■driven artificial intelligence systems—An introductory survey

Year: 2020 | Citations: 798 | Authors: Eirini Ntoutsi, P. Fafalios, U. Gadiraju, Vasileios Iosifidis, W. Nejdl

Abstract

Artificial Intelligence (AI)■based systems are widely employed nowadays to make decisions that have far■reaching impact on individuals and society. Their decisions might affect everyone, everywhere, and anytime, entailing concerns about potential human rights issues. Therefore, it is necessary to move beyond traditional AI algorithms optimized for predictive performance and embed ethical and legal principles in their design, training, and deployment to ensure social good while still benefiting from the huge potential of the AI technology. The goal of this survey is to provide a broad multidisciplinary overview of the area of bias in AI systems, focusing on technical challenges and solutions as well as to suggest new research directions towards approaches well■grounded in a legal frame. In this survey, we focus on data■driven AI, as a large part of AI is powered nowadays by (big) data and powerful machine learning algorithms. If otherwise not specified, we use the general term bias to describe problems related to the gathering or processing of data that might result in prejudiced decisions on the bases of demographic features such as race, sex, and so forth.